

WHAT IS CLAIMED IS:

- Sub  
Q1
1. A method for transforming words to unique numerical representations, comprising:  
receiving a text including multiple words; and  
transforming each of the received words into a unique numeral representation such that the transformed unique numerical representation does not result in multiple similar numerical representations, to avoid ambiguous prediction of meaning of the translated words in the received text.
2. The method of claim 1, wherein receiving the text comprises:  
receiving the text from a source selected from the group comprising a data base/data warehouse, a LAN/WAN network, the Internet, a voice recognition system, and a mobile/fixed phone.
3. The method of claim 1, further comprising:  
filtering the received words to extract one or more key-words; and  
morphologizing each of the filtered one or more key-words for base formatting based on similarities of fundamental characteristics in the one or more words
4. The method of claim 3, wherein the filtering the received words to extract the key-words comprises:  
filtering the received words to extract one or more key-words based on a specific criteria selected from the group comprising filtering to remove all words comprised of three or fewer letters, and filtering to remove rarely used words.

5. The method of claim 3, further comprising:  
inputting each of the transformed unique numerical representations for text mining applications such as automated email responses, automated text summarizations, and/or any other similar text mining application.

6. The method of claim 3, wherein the received text can be in any natural language.

7. The method of claim 1, wherein transforming each of the received words to a unique numerical representation, further comprises:

using an A to Z helix transformation function, wherein the A to Z helix transformation function comprises:

$$(W) = \sum_{k=0}^l \{(\beta_{l-k})n^{l-k} + (l-k)\}$$

wherein W is a unique number obtained for a word having a length of  $l+1$  letters, wherein the letters in the word W can be represented as  $\beta_l \beta_{(l-1)} \beta_{(l-2)} \dots \beta_0$ , and also wherein  $\beta_i$  represents the letter in the  $i^{\text{th}}$  location of the alphabet in a particular language having  $n$  distinct letters in the alphabet of the language (for example, in the English language,  $n$  being equal to 26).

8. A computer-implemented system for transforming words in a text to unique numerical representations, comprising:

a web server to receive the text including multiple words in a natural language;

a key-word extractor to extract one or more key-words from the received words;

a morphologizer to morphologize the extracted key-words based on similarities in fundamental characteristics of the extracted key-words; and

an analyzer to transform each of the morphologized words to a unique numerical representation such that the transformed unique numerical representation

does not result in multiple similar numerical representations, to avoid ambiguous prediction of meaning of the translated words in the received text.

9. The system of claim 8, wherein the key-word extractor extracts key-words based on a specific criteria selected from the group comprising filtering to remove all words including three or fewer letters in the received text, and filtering to remove rarely used words.

10. The system of claim 8, wherein the analyzer outputs the transformed words including unique numerical representations for use in text mining.

11. The system of claim 8, wherein the received text can be in any language.

12. The system of claim 8, wherein the analyzer further transforms each of the morphologized words to a unique numerical representation using an A to Z helix transformation function, wherein the A to Z helix transformation function comprises:

$$(W) = \sum_{k=0}^l \{(\beta_{l-k})n^{l-k} + (l-k)\}$$

wherein W is a unique number obtained for a word having a length of  $l+1$  letters, wherein the letters in the word W can be represented as  $\beta_l \beta_{(l-1)} \beta_{(l-2)} \dots \beta_0$ , and also wherein  $\beta_i$  represents the letter in the  $i^{\text{th}}$  location of the alphabet in a particular language having  $n$  distinct letters in the alphabet of the language (for example, in the English language,  $n$  being equal to 26).